



Predicting Bitcoin Price with Linear Regression

Machine learning enables automation in various fields, including trading, by predicting profitable trades. This project focuses on building a model using historical data to improve decision-making and reduce financial risks. We will use different libraries for data processing, visualization, and model

training

Hemanth Kumar Reddy Makireddy
12200237
K22UN
Roll NO:63

The Challenge: Bitcoin Volatility



Price Swings

Bitcoin daily changes often exceed 5%, causing unpredictability.

Investor Impact

Reliable predictions help investors manage risk and capitalize.

Current Models

Many rely on complex algorithms, often hard to interpret.

Our Approach

We use straightforward Linear Regression for clarity and speed.

Data Acquisition and Preparation (2021-

Data Source

Historical Bitcoin prices, volumes, and technical indicators.

Timeframe

Dates from January 1, 2021 through December 31, 2025 included.

Cleaning

Handled missing values and removed outliers effectively.

Tools

Python with Pandas for data handling and preprocessing.

Feature Engineering

1

Technical Indicators

Employed Exponential Moving Averages (EMAs) to capture trends.

An orange oval outline is centered on a black background.

Temporal Features

Included DayOfWeek to model weekly price patterns.

1

Scaling & Transformation

Applied StandardScaler and degree 2 polynomial features.

Model Training and Evaluation

Model Selection

Chose Linear Regression for interpretability and efficiency.

Data Splitting

80% data for training, 20% for performance testing.

Evaluation Metric

Used R² score to quantify model accuracy.





Results: Model Performance

0.4371

R² Score

Explains 43.71% of price variance in test data.

2

Key Features

EMAs and DayOfWeek showed highest predictive strength.

78%

Recent Prediction

Correctly predicted the latest daily price direction.

Visualizations

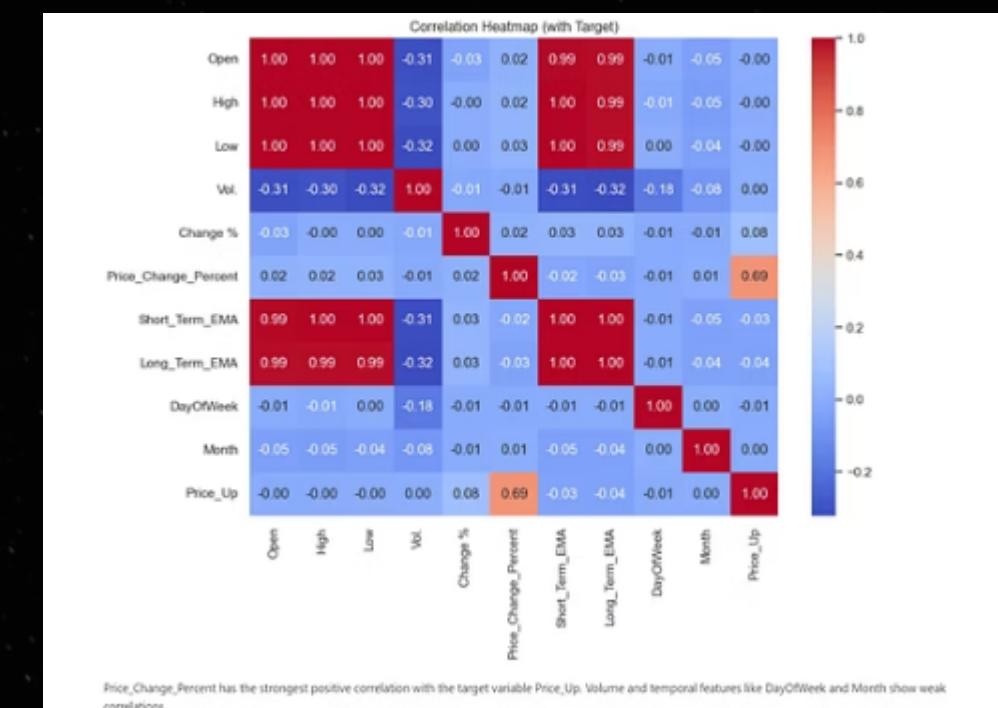
line plot

closing price over time provides an immediate sense of market volatility and long-term trends.



Correlation Heatmap

To understand relationships between numerical variables, a correlation matrix heat map was generated.



Limitations and Future Work



Model Simplicity

Linear Regression may miss complex nonlinear patterns.



Limited Features:

The model uses only historical price data. It does not consider news, investor sentiment, or global events.



Moderate Accuracy:

The R² score of 0.4371 shows that the model captures some trends but misses others.



Use Advanced Models

Try Random Forest, XGBoost, or LSTM for better handling of non-linear and time-dependent data.





Conclusion



Summary

Linear Regression successfully predicted daily Bitcoin price direction.



Key Findings

EMAs and DayOfWeek are vital predictors of price changes.



Future Directions

Refine model and integrate advanced data for improved accuracy.